Comparison Inequalities and Fastest-Mixing Markov Chains

(*Annals of Applied Probability*, to appear)

Jim Fill
(coauthor: Jonas Kahn, University of Lille)

Department of Applied Mathematics and Statistics
The Johns Hopkins University

July 21, 2012
8th Workshop on Markov Processes and Related Topics

FASTEST-MIXING MARKOV CHAINS: INTRO/SUMMARY

- FMMC problem: treated in a series of papers
  - Boyd, Diaconis, Xiao: *SIAM Rev.*, 2004
  - Sun, Boyd, Xiao, Diaconis: *SIAM Rev.*, 2006
  - Boyd, Diaconis, Sun, Xiao: *Amer. Math. Monthly*, 2006
  - Boyd, Diaconis, Parrilo, Xiao: *SIAM J. Optim.*, 2009
- given: finite graph $G = (V, E)$; probab. distn. $\pi > 0$ on V
- goal: Find the fastest-mixing reversible MC (FMMC) with stat. distn. π and transitions allowed only along the edges in E.
- very important problem because of MCMC [goal is (approx.) sampling from π, MC is constructed for efficient generation]
- their criterion for FMMC: minimize SLEM
- They find the FMMC using semidefinite programming.
- related work: Roch, *Electron. Comm. Probab.*, 2005

FMMC on a path

- Most of the results in the series of papers are numerical, but there are some analytical results, incl. for FMMC on a path (we'll call this the path problem).
- has application to load balancing for a network of processors (Diekmann, Muthukrishnan, and Nayakkankuppam, *Lecture Notes in Computer Science*, 1997)
- G = path on $V = \{0, \ldots, n\}$ with a self-loop at each vertex
- π is uniform on V
- It is proved that the FMMC (in terms of SLEM) has transition probability $p(i, i+1) = p(i+1, i) = 1/2$ along each edge and $p(i, i) \equiv 0$ except that $p(0, 0) = 1/2 = p(n, n)$.
- We call this the *uniform* chain (for short: UC) $U = (U_t)_{t=0,1,\ldots}$.

True fastest mixing

- Various measures of mixing time for a MC can indeed be bounded using the SLEM, which provides the asymptotic exponential rate of convergence to stationarity.
- But the SLEM provides only a surrogate for **true** measures of discrepancy from stationarity, such as total variation (TV) distance, separation (sep), and L^2-distance.
- For the path problem, Diaconis wondered whether the uniform chain might in fact minimize such distances after any given number of steps, when all chains considered start at 0.
- We show: The UC is **truly** FM in a wide variety of senses.

Majorization and fastest mixing

- What we show, precisely, is that, for any B&D chain X having symmetric transition kernel on the path and initial state 0, and for any $t \geq 0$, the pmf π_t of X_t majorizes the pmf σ_t of U_t.
- We show using this that four examples of discrepancy from uniformity that are larger for X_t than for U_t are
  - (i) $L^p(\pi)$-distance for any $1 \leq p \leq \infty$ (including TV & L^2);
  - (ii) separation;
  - (iii) Hellinger distance;
  - (iv) Kullback–Leibler divergence.
- Our new (and simple!) technique used to prove that π_t majorizes σ_t is quite general: **comparison inequalities (CIs)**.
- We show that if two Markov semigroups satisfy a certain CI at time 1, then they satisfy the same CI at all times t.
- We also show how the CI can be used to compare mixing times—in a variety of senses—for the chains with the given semigroups.

The CI-approach

- We show that, in the context of the path-problem, if one restricts either (i) to monotone chains, or (ii) to even times, then the UC satisfies a favorable CI in comparison with any other chain in the class considered.
- Delicate arguments (needed except for L^2-distance) specific to the path-problem allow us to remove the parity restriction.
- Further, comparisons between chains—even time-inhomogeneous ones—other than the UC can be carried out with our CI method by limiting attention either to monotone kernels or to two-step kernels.
- Indeed, our CI-approach rather generally provides a new tool for the notoriously difficult analysis of time-inhomogeneous chains, whose nascent quantitative theory has been advanced impressively in recent work of Saloff-Coste and Zúñiga.

Comparison inequalities: two other applications

1. We generalize our path-problem result: Let π be a log-concave pmf on $\mathcal{X} = \{0, \ldots, n\}$. Among all *monotone* B&D kernels K, we identify the fastest to mix (again, in a variety of senses). The fastest K reduces to the UC kernel when π is uniform.
2. We show how CIs can recover and extend (among other ways, to certain card-shuffling chains) a Peres–Winkler result about slowing down mixing by skipping ("censoring") updates of monotone spin systems. (This is an example of CIs applied to time-inhomogeneous chains.)

END OF SUMMARY

COMPARISON INEQUALITIES: set-up

Let's set up:

- given: a pmf $\pi > 0$ on a finite partially ordered state space $\mathcal{X}$
- the usual $L^2(\pi)$ inner product :

$$\langle f, g \rangle \equiv \langle f, g \rangle_\pi := \sum_{i \in \mathcal{X}} \pi(i) f(i) g(i)$$

- the $L^2(\pi)$-adjoint (aka time-reversal) of a kernel K:

$$K^*(i,j) \equiv \pi(j) K(j,i) / \pi(i)$$

- reversibility $\equiv$ self-adjointness
- $\mathcal{K} := \{$Markov kernels on $\mathcal{X}$ with stat. distn. $\pi\}$
- $\mathcal{M} := \{$nonnegative non-increasing functions on $\mathcal{X}\}$
- $\mathcal{S} := \{K \in \mathcal{K} : K$ is stochastically monotone$\}$
  (Note: K is said to be SM if $Kf \in \mathcal{M}$ for every $f \in \mathcal{M}$.)
  (Note: The identity kernel I belongs to $\mathcal{S}$, regardless of π.)

Comparison inequalities: definition

Definition of *comparison inequality* (CI) relation $\preceq$ on $\mathcal{K}$:
We write $K \preceq L$ if $\langle Kf, g\rangle \le \langle Lf, g\rangle$ for every $f, g \in \mathcal{M}$.

Observe: $K \preceq L$ iff the time-reversals K^* and L^* satisfy $K^* \preceq L^*$.

Remark

(a) Indicators of down-sets are enough to establish a CI.

(b) There is an important existing notion of stochastic ordering for Markov kernels on $\mathcal{X}$: We say that $L \le_{\text{st}} K$ if $Kf \le Lf$ entrywise for all $f \in \mathcal{M}$. It is clear that $L \le_{\text{st}} K$ implies $K \preceq L$ when K and L belong to $\mathcal{S}$. But in all our examples where we prove a comparison inequality, we do not have stochastic ordering. This will typically be the case for interesting examples, since the requirement for distinct $K, L \in \mathcal{S}$ to have the same stationary distribution makes it difficult (though not impossible) to have $L \le_{\text{st}} K$.

Comparison inequalities: give a partial order on $\mathcal{K}$

Remark

The relation $\preceq$ defines a partial order on $\mathcal{K}$. Indeed:

- Reflexivity and transitivity are immediate.
- Antisymmetry follows because one can build a basis for functions on $\mathcal{X}$ from elements f of $\mathcal{M}$, namely, the indicators of principal down-sets (i.e., down-sets of the form $\langle x \rangle := \{y : y \leq x\}$ with $x \in \mathcal{X}$).

Comparison inequalities: basic properties of $\preceq$ on $\mathcal{K}$

- Claim: The CI relation $\preceq$ on $\mathcal{K}$ is preserved under passages to limits, mixtures, and direct sums. (See the next Proposition.)
- Note: The class $\mathcal{S}$ is closed under passages to limits and mixtures, and also under (finite) products, but not under general direct sums as in part (c) of the next Proposition.

Proposition

(a) If $K_t \preceq L_t$ for every t and $K_t \to K$ and $L_t \to L$, then $K \preceq L$.

(b) If $K_t \preceq L_t$ for $t = 0, 1$ and $0 \le \lambda \le 1$, then

$$(1-\lambda)K_0 + \lambda K_1 \preceq (1-\lambda)L_0 + \lambda L_1.$$

(c) Partition $\mathcal{X}$ arbitrarily into subsets $\mathcal{X}_0$ and $\mathcal{X}_1$, and let each $\mathcal{X}_i$ inherit its p.o. and stat. distn. from $\mathcal{X}$. For $i = 0, 1$, suppose $K_i \preceq L_i$ on $\mathcal{X}_i$. Define $K := K_0 \oplus K_1$ & $L := L_0 \oplus L_1$. Then $K \preceq L$.

Comparison inequalities: preservation under product

Our main result for the CI relation $\preceq$:

Proposition (CIs: preservation under product)

Let $K_1, \ldots, K_t$ and $L_1, \ldots, L_t$ be reversible kernels all belonging to $\mathcal{S}$, and suppose that $K_s \preceq L_s$ for $s = 1, \ldots, t$. Then the product kernels $K_1 \cdots K_t$ and $L_1 \cdots L_t$ (and their time-reversals) belong to $\mathcal{S}$, and $K_1 \cdots K_t \preceq L_1 \cdots L_t$.

Application to time-homogeneous chains:

Corollary

If $K, L \in \mathcal{S}$ are reversible and $K \preceq L$, then for every t we have $K^t, L^t \in \mathcal{S}$ and $K^t \preceq L^t$.

CI-technique: applicability

Remark

As we shall see from examples, the applicability of our new CI-technique is limited

(i) by the monotonicity requirement for membership in $\mathcal{S}$ and

(ii) by the limited extent to which $\mathcal{S}$ is ordered by $\preceq$.

But restriction (i) in the choice of kernel has the payoff (among others) that the perfect simulation algorithms

- Coupling From The Past of Propp and Wilson (*Random Structures Algorithms*, 1996) and
- FMMR (F, *Ann. Appl. Probab.*, 1998; F, Machida, Murdoch, and Rosenthal, *Random Structures Algorithms*, 2000)

can often be run efficiently for monotone chains.

CONSEQUENCES OF COMPARISON INEQUALITIES

Establishment of CIs leads to comparisons of mixing speed.

In the simple case of time-homogeneous reversible chains with "nice" initial distributions, this is the reason why:

1. A comparison inequality implies stochastic domination.
2. Domination implies an inequality in mixing speed.

Definition of domination

Let (Y_t) and (Z_t) be stochastic processes with the same finite partially ordered state space. If for every t we have $Y_t \geq Z_t$ stochastically, i.e.,

$$\mathbf{P}(Y_t \in D) \leq \mathbf{P}(Z_t \in D) \text{ for every down-set } D,$$

then we say that Y *dominates* Z.

Comparison inequalities and domination

Proposition (relating comparison inequalities and domination)

Suppose that $K, L \in \mathcal{S}$ are reversible and satisfy $K \preceq L$. If Y and Z are chains

(i) started in a common pmf $\hat{\pi}$ such that $\hat{\pi}/\pi$ is non-increasing &

(ii) having respective kernels K and L,

then Y dominates Z.

Proof.

By preservation of CI under product, for every t we have $K^t, L^t \in \mathcal{S}$ and $K^t \preceq L^t$. The desired result now follows easily. □

next slide: Domination is quite useful for comparing mixing times in at least three standard senses.

TV, separation, and L^2-distance

If d is a measure of discrepancy from stat., then in the following thm. we write "Y mixes faster in d than does Z" for the strong assertion that at every time t we have d smaller for Y than for Z.

Theorem

Consider reversible Markov chains Y and Z with common finite p.o. state space $\mathcal{X}$, common init. distn. $\hat{\pi}$, and common stat. distn. π. Assume that $\hat{\pi}/\pi$ is non-increasing.

(a) **[total variation distance]** *Suppose that Y dominates Z and that the kernel of Y belongs to $\mathcal{S}$. Then Y mixes faster in TV than does Z.*

(b) **[separation]** *Same hypotheses as in (a). Then Y mixes faster in separation than does Z; equivalently, any fastest strong stationary time for Y is stochastically smaller (i.e., faster) than any strong stationary time for Z.*

TV, separation, and L^2-distance (continued)

Theorem (continued)

(c) **[L^2-distance]** *Suppose that the two-step chain* (Y_{2t}) *dominates* (Z_{2t}) *and has kernel in* $\mathcal{S}$. *Then* Y *mixes faster in* L^2 *than does* Z.

Remark [concerning eigenvalues]

(a) If K and L are ergodic reversible kernels in $\mathcal{S}$ (with a common stat. distn. π) and we have the comparison inequality $K \preceq L$, then the SLEM for K is no larger than the SLEM for L.

(b) There are several existing std. techniques for comparing mixing times of MCs, such as the celebrated eigenvalues-comparison technique of Diaconis and Saloff-Coste (*Ann. Appl. Probab.*, 1993), but none give conclusions as strong as ours. On the other hand, comparison of eigenvalues requires verifying far fewer assumptions than ours, so our new technique is much less generally applicable.

Other distances via majorization

- Given vectors v and w in $\mathbf{R}^N$, we say that v majorizes w if
  (i) for each $k = 1, \ldots, N$, the sum of the k largest entries of w is at least the corresponding sum for v, and
  (ii) equality holds when $k = N$.
- A function ϕ with domain $D \subseteq \mathbf{R}^N$ is Schur-convex on D if $\phi(v) \geq \phi(w)$ whenever $v, w \in D$ and v majorizes w.
- Thus, given any two pmfs ρ_1 and ρ_2 on $\mathcal{X}$, if ρ_1 majorizes ρ_2, then for any Schur-convex function ϕ on the unit simplex (i.e., the space of pmfs) we have $\phi(\rho_1) \geq \phi(\rho_2)$.
- We give six examples below where a conclusion of the form "ρ_2 is closer to π than is ρ_1" follows from an inequality $\phi(\rho_1) \geq \phi(\rho_2)$ for a Schur-convex function ϕ (all of which follow in turn from majorization of ρ_2 by ρ_1).

Other distances via majorization

The next proposition describes one important case where we have majorization and hence can extend the conclusions "Y mixes faster in d than does Z" to other measures of discrepancy d (when the additional hypothesis that π is non-increasing is strengthened further to the assumption that π is uniform).

Proposition

Suppose that $K, L \in \mathcal{S}$ are reversible kernels on a common finite p.o. state space $\mathcal{X}$ and satisfy $K \preceq L$, and that their common stat. distn. π is non-increasing. If Y and Z are chains

(i) *started in a common pmf $\hat{\pi}$ such that $\hat{\pi}/\pi$ is non-increasing &*

(ii) *having respective kernels K and L,*

then, for all t, the pmf π_t of Z_t majorizes the pmf σ_t of Y_t.

Other distances via majorization: examples

When π is uniform in the preceding prop., then Y mixes faster than does Z in, as examples, the following six senses (where here we have written the discrepancy from $\pi = \text{unif}$ for a generic pmf ρ):

(i) L^p-distance

$$\left[\sum_i \pi(i) \left|\frac{\rho(i)}{\pi(i)} - 1\right|^p\right]^{1/p},$$

for any $1 \le p < \infty$;

(ii) L^∞-distance

$$\max_i \left|\frac{\rho(i)}{\pi(i)} - 1\right|,$$

also called relative pointwise distance;

(iii) separation

$$\max_i \left[1 - \frac{\rho(i)}{\pi(i)}\right];$$

Other distances via majorization: examples

(iv) Hellinger distance

$$\frac{1}{2}\sum_i \pi(i)\left[\sqrt{\frac{\rho(i)}{\pi(i)}}-1\right]^2;$$

(v) the Kullback–Leibler divergence

$$D_{KL}(\pi\|\rho) = -\sum_i \pi(i)\ln\left[\frac{\rho(i)}{\pi(i)}\right];$$

(vi) the Kullback–Leibler divergence

$$D_{KL}(\rho\|\pi) = \sum_i \rho(i)\ln\left[\frac{\rho(i)}{\pi(i)}\right].$$

The L^2-distance considered earlier is the special case $p = 2$ of ex. (i) here, and TV distance amounts to the special case $p = 1$.

FASTEST MIXING ON A PATH

We now specialize to the path-problem.

- Let K be any symmetric B&D kernel on the path $\{0, 1, \ldots, n\}$.
- Note!: $K \in \mathcal{S}$ iff $K(i, i+1) + K(i+1, i) \leq 1$ for $0 \leq i \leq n-1$.
- Assume K is symmetric and denote $K(i, i+1) = K(i+1, i)$ by p_i [except: $K(0,0) = 1 - p_0$ and $K(n,n) = 1 - p_{n-1}$].
- For example, when $n = 3$ we have

$$K = \begin{bmatrix} 1-p_0 & p_0 & 0 & 0 \\ p_0 & 1-p_0-p_1 & p_1 & 0 \\ 0 & p_1 & 1-p_1-p_2 & p_2 \\ 0 & 0 & p_2 & 1-p_2 \end{bmatrix}.$$

We show that if one restricts attention either (i) to monotone chains, or (ii) to even times, then the UC U with kernel K_0 where $p_i \equiv 1/2$ satisfies a favorable CI in comparison with the general K-chain.

Fastest mixing on a path

Before we separate into the two cases (i) and (ii) for the path-problem, let's note that if f is the indicator of the down-set $\{0, 1, \ldots, \ell\}$, then Kf satisfies

$$(Kf)_j = \begin{cases} 1 & \text{if } 0 \le j \le \ell - 1 \\ 1 - p_\ell & \text{if } j = \ell \\ p_\ell & \text{if } j = \ell + 1 \\ 0 & \text{otherwise} \end{cases}$$

(with $p_n = 0$); hence if g is the indicator of the down-set $\{0, 1, \ldots, m\}$, then

$$\langle Kf, g \rangle = \frac{1}{n+1} \times \begin{cases} m + 1 & \text{if } 0 \le m \le \ell - 1 \\ \ell + 1 - p_\ell & \text{if } m = \ell \\ \ell + 1 & \text{if } \ell + 1 \le m \le n. \end{cases}$$

(i) Restriction to monotone chains

- Our symmetric kernel K is monotone if and only if $p_i \leq 1/2$ for $i = 0, \dots, n-1$.
- Among all such choices, it is clear that $\langle Kf, g \rangle$ from the preceding slide is minimized when $K = K_0$.
- It therefore follows that $K_0 \preceq K$ and hence that K_0 is fastest-mixing in several senses.
- In fact, we see that monotone symmetric B&D kernels K are monotonically decreasing in the partial order $\preceq$ with respect to each p_i.

(ii) Restriction to even times

- In the present setting of symmetric birth-and-death kernel, note that our restriction (simply to ensure that K is a kernel) on the values $p_i > 0$ is that $p_i + p_{i+1} \leq 1$ for $i = 0, \ldots, n-1$.
- It is then routine to check that K^2 is (like K) reversible and (perhaps unlike K) monotone. Indeed, if f is the indicator of the down-set $\{0, 1, \ldots, \ell\}$, then $K^2 f$ satisfies

$$(K^2 f)_j = \begin{cases} 1 & \text{if } 0 \leq j \leq \ell - 2 \\ 1 - p_{\ell-1} p_\ell & \text{if } j = \ell - 1 \\ 1 - 2p_\ell + 2p_\ell^2 + p_{\ell-1} p_\ell & \text{if } j = \ell \\ 2p_\ell - 2p_\ell^2 - p_\ell p_{\ell+1} & \text{if } j = \ell + 1 \\ p_\ell p_{\ell+1} & \text{if } j = \ell + 2 \\ 0 & \text{otherwise,} \end{cases}$$

which is easily checked to be non-increasing in j.

(ii) Restriction to even times

- Suppose now that g is the indicator of the down-set $\{0, 1, \ldots, m\}$. Then using $K^2 f$ from the preceding slide we can calculate, and subsequently minimize over the allowable choices of $p_0, \ldots, p_{n-1}$, the quantity $\langle K^2 f, g \rangle$.
- The minimum is achieved by the UC ($p_i \equiv 1/2$).
- It therefore follows that $K_0^2 \preceq K^2$ and hence that K_0^2 is fastest-mixing in several senses.
- Specifically:

  for all even t, the pmf π_t of X_t majorizes the pmf σ_t of U_t

  if X and U have respective kernels K and K_0 and common non-increasing initial pmf $\hat{\pi}$.
- Further, when we consider all symmetric B&D chains started in state 0, it follows that the UC is fastest-mixing in L^2 (without the need to restrict to even times, nor to monotone chains).

(ii) Restriction to even times; removal of parity restriction

Remark

We see more generally that if K and $\widetilde{K}$ are two symmetric B&D kernels and for every i we have

$$\left|p_i - \tfrac{1}{2}\right| \geq \left|\tilde{p}_i - \tfrac{1}{2}\right| \quad \text{and} \quad p_i p_{i+1} \leq \tilde{p}_i \tilde{p}_{i+1},$$

then $\widetilde{K}^2 \preceq K^2$.

Delicate arguments (needed except for L^2-distance) specific to the path-problem allow us to remove the parity restriction:

Theorem

Let X be a B&D chain on $\mathcal{X} = \{0, 1, \dots, n\}$ and symmetric kernel, and let U be the UC. Suppose that both chains start at 0, and let π_t (resp., σ_t) denote the pmf of X_t (resp., U_t). Then

$$\pi_t \textit{ majorizes } \sigma_t \textit{ for all } t = 0, 1, 2 \dots.$$

FMMC for path-problem: remarks

Remark

(a) The multiset of values $\{\mathbf{P}_i(U_t = j) : j \in \{0, \ldots, n\}\}$ for the uniform chain U started in state i does not depend on $i \in \{0, \ldots, n\}$; therefore, the uniform chain minimizes various distances from stationarity (including all six listed earlier) not only when the starting state is 0 but in the worst case over all starting states (and indeed over all starting distributions).

(b) The SLEM is an asymptotic measure (in the worst case over starting states) of distance from stationarity. Accordingly, by remark (a), the uniform chain minimizes SLEM among all symmetric B&D chains. Thus we recover the main result of Boyd, Diaconis, Sun, and Xiao (*Amer. Math. Monthly*, 2006).

FASTEST-MIXING MONOTONE B&D CHAIN: fixed π

Theorem

Let π be log-concave on $\mathcal{X} = \{0, \dots, n\}$. Let K_π have (death, hold, birth) probabilities (q_i, r_i, p_i) given by

$$q_i = \frac{\pi_{i-1}}{\pi_{i-1} + \pi_i}, \quad r_i = \frac{\pi_i^2 - \pi_{i-1}\pi_{i+1}}{(\pi_{i-1} + \pi_i)(\pi_i + \pi_{i+1})}, \quad p_i = \frac{\pi_{i+1}}{\pi_i + \pi_{i+1}},$$

with $\pi_{-1} := 0$ and $\pi_{n+1} := 0$. Then K_π is a monotone B&D kernel with stat. distn. π, and $K_\pi \preceq K$ for any such kernel K.

Remark

More generally, the kernels $K \in \mathcal{S}$ are non-increasing (in $\preceq$) in each p_i, and $p_i = \pi_{i+1}/(\pi_i + \pi_{i+1})$ maximizes p_i subject to the monotonicity constraint.

Fastest-mixing monotone B&D chains: random walks

Here is an example of a fastest-mixing monotone B&D chain:

- Suppose that the stationary pmf is proportional to $\pi_i \equiv \rho^i$, i.e., is either truncated geometric (if $\rho < 1$) or its reverse (if $\rho > 1$) or uniform (if $\rho = 1$).
- Then the kernel K_π corresponds to biased random walk:

$$q_i \equiv q := \frac{1}{1+\rho}, \qquad r_i \equiv 0, \qquad p_i \equiv p := \frac{\rho}{1+\rho},$$

with the following endpoint exceptions, of course:

$$q_0 = 0,\ r_0 = q,\ r_n = p,\ p_n = 0.$$

Slowest FMMC: the uniform chain

Theorem

Among the fastest-mixing monotone B&D chains (kernel $= K_\pi$) with initial state 0 and log-concave stationary pmf π, the uniform chain is slowest to mix in separation.

Here is a broad two-step outline of our proof:

1. We show (using the strong stationary duality theory of Diaconis and F (*Ann. Appl. Probab.*, 1991) that the chain with kernel K_π mixes faster in separation than does the biased random walk with ρ set to ρ_{i_0}, where

$$\rho_i := \pi_{i+1}/\pi_i \quad (i = 0, \ldots, n-1),$$

   and $i = i_0$ minimizes $|\ln \rho_i|$.
2. The biased random walks are monotonically slower to mix in separation as $\min\{\rho, \rho^{-1}\}$ increases.

CAN EXTRA UPDATES DELAY MIXING?

- Can extra updates delay mixing? This question is the title of a paper by Peres and Winkler (preprint, 2011).
- Peres and Winkler show that the answer is no, for TV, in the setting of monotone spin systems, generalized by replacing the set of spins $\{0, 1\}$ by any linearly ordered set. (We review relevant terminology below.)
- See also Holroyd (preprint, 2011) for counterexamples.
- We recapture and extend their result using CIs by showing that $K_v \preceq I$ for any kernel K_v that updates a single site v, i.e., that replacing K_v by the identity kernel only slows mixing (when the initial pmf has non-increasing ratio with respect to the stationary pmf)—because then, noting reversibility and stoch. mono. of each K_v, for any $v_1, \ldots, v_t$ the product $K_{v_1} \cdots K_{v_t}$ increases in $\preceq$ by deletion of any K_{v_i}.

Positive correlations

- The CI $K_v \preceq I$ holds in the more general setting of a poset of "spins", subject to the following restriction: Starting with distribution π and a site v and conditioning on the spins at all sites other than v, the conditional law of the spin at v should have positive correlations.
- Recall that a pmf π on a finite partially ordered set $\mathcal{X}$ is said to have positive correlations (PCs) if

$$\langle f, g \rangle \geq \langle f, 1 \rangle \langle g, 1 \rangle$$

for every $f, g \in \mathcal{M}$, and that if $\mathcal{X}$ is *linearly* ordered then (by "Chebyshev's other inequality") *all* probability measures have PCs. The connection with comparison inequalities is the following simple lemma (note that both K_π and I belong to $\mathcal{S}$).

Positive correlations

Lemma

A pmf π on a finite partially ordered set $\mathcal{X}$ has PCs if and only if $K_\pi \preceq I$, where K_π is the trivial kernel that jumps in one step to π and I is the identity kernel.

Proof.

Since for any f and g we have

$$\langle K_\pi f, g \rangle = \langle \langle f, 1 \rangle 1, g \rangle = \langle f, 1 \rangle \langle g, 1 \rangle$$

and $\langle If, g \rangle = \langle f, g \rangle$, the lemma is proved. □

Positive correlations

Proposition

Let π be a pmf on a finite poset. Partition $\mathcal{X}$, suppose that a given kernel K on $\mathcal{X}$ is a direct sum of trivial kernels K_i (as in the preceding lemma) on the cells of the partition, and suppose that π conditioned to each cell has PCs. Then $K \preceq I$.

Proof.

Simply combine the preceding lemma with preservation of $\preceq$ under direct product. □

Monotone spin systems

Our setting is the following:

- given: finite graph $G = (V, E)$; finite poset S of "spin values"
- A spin config. is an assignment of spins to vertices (sites).
- Our state space is the set $\mathcal{X}$ of all spin configurations.
- given: a pmf π on $\mathcal{X}$ that is monotone in the sense that, when we start with π and any site v and condition on the spins at all sites other than v, the conditional law of the spin at v is monotone in the conditioning spins.

We recover and (modestly) extend the Peres–Winkler result by means of the following theorem, which

(i) allows somewhat more general S and

(ii) encompasses separation and L^2-distance as well as TV.

Monotone spin systems

Theorem

Fix a site v, and suppose that the conditional distributions discussed above all have PCs. Let K_v be the (stochastically monotone) Markov kernel for update at site v according to the conditional distributions discussed. Then we have the comparison inequality $K_v \preceq I$.

Remark (random vs. systematic site updates)

It follows for monotone spin systems with (say) linearly ordered S that, when the chains start from a common pmf having non-increasing ratio relative to π, the "systematic site updates" chain with kernel $K_{\mathrm{syst}} := K_{v_1} \cdots K_{v_\nu}$ (for any ordering $v_1, \ldots, v_\nu$ of the sites $v \in V$) mixes faster in TV, sep, and L^2 than does the "random site updates" chain with kernel $K_{\mathrm{rand}} := \sum_{v \in V} p_v K_v$ [for any pmf $\mathbf{p} = (p_v)_{v \in V}$ on V].

Monotone spin systems

Remark (random vs. systematic site updates, continued)

It is important to keep in mind here that one "sweep" of the sites using K_{syst} is counted as only one Markov-chain step.

There is a very weak ordering in the opposite direction: $K^{\nu}_{\mathrm{rand}} \preceq pK_{\mathrm{syst}} + (1-p)I$, with $p := \prod_{v \in V} p_v$.

Extra updates don't delay mixing: card-shuffling

The following card-shuffling Markov chain is another example where CIs can be used to show that extra updates do not delay mixing.

- has been studied quite a bit [see Benjamini, Berger, Hoffman, and Mossel (*Trans. Amer. Math. Soc.*, 2005) and references therein] in the time-homogeneous "random updates" case where update positions are chosen independently and uniformly
- state space $\mathcal{X} = \{\text{permuts. of } \{1, \dots, n\}\}$; param. $p \in (0, 1)$
- Given $i \in \{1, \dots, n-1\}$, we can update adjacent positions i and $i+1$ by sorting the two cards in those positions wp p and "anti-sorting" them wp $1-p$. Call this update kernel K_i.
- can check: Each K_i is
  - (i) reversible wrt $\pi(x)$ proportional to $[(1-p)/p]^{\text{inv}(x)}$, and
  - (ii) stochastically monotone with respect to the Bruhat order on $\mathcal{X}$ (defined so that $x \leq y$ if y can be obtained from x by a sequence of anti-sorts of *not necessarily adjacent* cards).

Extra updates don't delay mixing: card-shuffling

Theorem

Fix a position $i \in \{1, \ldots, n-1\}$, and let K_i be the Markov kernel for update of positions i and $i+1$ as discussed above. Then we have the comparison inequality $K_i \preceq I$.

The key is that the cells of the relevant partition of $\mathcal{X}$ now consist of only two permutations each and are each clearly linearly ordered, therefore having PCs.

A final example

In a specific setting (linearly ordered state space and uniform stationary distribution) we have $K \preceq I$ quite generally:

Theorem

Let $\mathcal{X}$ be a linearly ordered state space. If K is doubly stochastic, then $K \preceq I$ (with respect to uniform π).

Remark

(a) Inserting a mono. symmetric kernel in a list of such kernels to be applied never slows mixing when the initial pmf is non-increasing.

(b) If "linearly ordered" is relaxed to "partially ordered" in the theorem, the result is not generally true, even for monotone K. This follows from the lemma characterizing PCs as trivial $\preceq I$, since there are posets for which the uniform distn. does not have PCs.

References

Itai Benjamini, Noam Berger, Christopher Hoffman, and Elchanan Mossel.
Mixing times of the biased card shuffling and the asymmetric exclusion process.
Trans. Amer. Math. Soc., 357(8):3013–3029 (electronic), 2005.

Stephen Boyd, Persi Diaconis, Pablo Parrilo, and Lin Xiao.
Fastest mixing Markov chain on graphs with symmetries.
SIAM J. Optim., 20(2):792–819, 2009.

Stephen Boyd, Persi Diaconis, Jun Sun, and Lin Xiao.
Fastest mixing Markov chain on a path.
Amer. Math. Monthly, 113(1):70–74, 2006.

References (continued)

Stephen Boyd, Persi Diaconis, and Lin Xiao.
Fastest mixing Markov chain on a graph.
SIAM Rev., 46(4):667–689 (electronic), 2004.

Persi Diaconis and James Allen Fill.
Strong stationary times via a new form of duality.
Ann. Probab., 18(4):1483–1522, 1990.

Persi Diaconis and Laurent Saloff-Coste.
Comparison theorems for reversible Markov chains.
Ann. Appl. Probab., 3(3):696–730, 1993.

References (continued)

Ralf Diekmann, S. Muthukrishnan, and Madhu V. Nayakkankuppam.
Engineering diffusive load balancing algorithms using experiments.
In *Lecture Notes in Computer Science*, volume 1253, pages 111–122. Springer, 1997.

James Allen Fill.
An interruptible algorithm for perfect sampling via Markov chains.
Ann. Appl. Probab., 8(1):131–162, 1998.

References (continued)

James Allen Fill, Motoya Machida, Duncan J. Murdoch, and Jeffrey S. Rosenthal.
Extension of Fill's perfect rejection sampling algorithm to general chains.
Random Structures Algorithms, 17(3-4):290–316, 2000. Special issue: Proceedings of the Ninth International Conference "Random Structures and Algorithms" (Poznan, 1999).

Alexander E. Holroyd.
Some circumstances where extra updates can delay mixing, 2011.
Preprint available: `arxiv::1101.4690v1`.

References (continued)

Yuval Peres and Peter Winkler.
Can extra updates delay mixing?, 2011.
Preprint, arXiv:1112.0603v1 [math.PR].

James Gary Propp and David Bruce Wilson.
Exact sampling with coupled Markov chains and applications to statistical mechanics.
Random Structures Algorithms, 9:223–252, 1996.

Sébastien Roch.
Bounding fastest mixing.
Electron. Comm. Probab., 10:282–296 (electronic), 2005.

References (continued)

L. Saloff-Coste and J. Zúñiga.
Convergence of some time inhomogeneous Markov chains via spectral techniques.
Stochastic Process. Appl., 117(8):961–979, 2007.

L. Saloff-Coste and J. Zúñiga.
Merging for time inhomogeneous finite Markov chains. I. Singular values and stability.
Electron. J. Probab., 14:1456–1494, 2009.

L. Saloff-Coste and J. Zúñiga.
Time inhomogeneous markov chains with wave like behavior.
Ann. Appl. Probab., 20(5):1831–1853, 2010.

References (continued)

L. Saloff-Coste and J. Zúñiga.
Merging for inhomogeneous finite markov chains, part ii: Nash and log-sobolev inequalities.
Ann. Probab., to appear.

Jun Sun, Stephen Boyd, Lin Xiao, and Persi Diaconis.
The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem.
SIAM Rev., 48(4):681–699 (electronic), 2006.